METHOD AND APPARATUS FOR PERFORMING OPTICAL CHARACTER RECOGNITION (OCR) AND TEXT STITCHING

The Field of the Invention

5   This invention relates generally to optical character recognition (OCR). This invention relates more particularly to a method and apparatus for performing optical character recognition and text stitching.

Background of the Invention

10   Many electronic devices, such as cellular telephones and personal digital assistants (PDAs), have the need for a digital camera to be included in the design. Such combined devices have been manufactured. The digital cameras for such combined devices are designed for general photography use. These cameras can also be used to capture images of printed text, such as the text on

15   business cards. If OCR is performed on a digital image of text, a text file may be generated.

          OCR requires high definition images. For some documents, several hundred thousand pixels or more may be required to obtain the desired recognition accuracy. However, some digital cameras, such as some digital

20   cameras for cell phones, may only have a small number of pixels (e.g., 352 x 288). In such limited-pixel systems, only a small portion of a document can be imaged at a high enough resolution for OCR. Multiple images of a document can be "stitched" together to create a larger image with more pixels. Then, OCR can be performed on the larger image. But it is computationally intensive to

25   perform such image stitching, and the lens distortion of multiple images makes stitching very difficult, if not impossible, in some cases.

          It would be desirable to provide a more accurate and less computationally intensive system and method for converting digital images of a document into an electronic text file.

30

Summary of the Invention

One form of the present invention provides a method of generating an electronic text file from a paper-based document that includes a plurality of characters. A plurality of partially overlapping digital images of the document are captured. Optical character recognition is performed on each one of the plurality of captured digital images, thereby generating a corresponding plurality of electronic text files. Each one of the electronic text files includes a portion of the plurality of characters in the document. The plurality of electronic text files are compared with one another to identify characters that are in common between the electronic text files. The plurality of electronic text files are combined into a combined text file based on the comparison. The combined text file includes the plurality of characters in the document.

Brief Description of the Drawings

Figure 1A is a diagram illustrating a simplified front view of a prior art digital camera.

Figure 1B is a diagram illustrating a simplified rear view of the digital camera shown in Figure 1A.

Figure 2 is a block diagram illustrating major components of the digital camera shown in Figures 1A and 1B.

Figure 3 is a block diagram illustrating major components of a digital camera configured to perform OCR and text stitching according to one embodiment of the present invention.

Figure 4 is a diagram illustrating four overlapping image frames of text information.

Figure 5 is a flow diagram illustrating an OCR and text stitching algorithm according to one embodiment of the present invention.

Figure 6A is a diagram illustrating a front side of a combined cellular telephone and digital camera device configured to perform OCR and text stitching according to one embodiment of the present invention.

Figure 6B is a diagram illustrating a back side of the combined cellular telephone and digital camera device shown in Figure 6A.

Description of the Preferred Embodiments

5   In the following detailed description of the preferred embodiments, reference is made to the accompanying drawings, which form a part hereof, and in which is shown by way of illustration specific embodiments in which the invention may be practiced. It is to be understood that other embodiments may be utilized and structural or logical changes may be made without departing

10   from the scope of the present invention. The following detailed description, therefore, is not to be taken in a limiting sense, and the scope of the present invention is defined by the appended claims.

Figure 1A is a diagram illustrating a simplified front view of a prior art digital camera 100. Figure 1B is a diagram illustrating a simplified rear view of

15   the digital camera 100 shown in Figure 1A. As shown in Figures 1A and 1B, camera 100 includes shutter button 102, optical viewfinder 104, flash 106, lens 108, liquid crystal display (LCD) 112, and user input device 114. User input device 114 includes buttons 114A-114C. User input device 114 allows a user to enter data and select various camera options.

20   In operation, a user looks through optical viewfinder 104 or at LCD 112 and positions camera 100 to capture a desired image. When camera 100 is in position, the user presses shutter button 102 to capture the desired image. An optical image is focused by lens 108 onto image sensor 200 (shown in Figure 2), which generates pixel data that is representative of the optical image. Captured

25   images are displayed on display 112. Flash 106 is used to illuminate an area to capture images in low-light conditions.

Figure 2 is a block diagram illustrating major components of digital camera 100. Camera 100 includes lens 108, image sensor 200, shutter controller 204, processor 206, memory 208, input/output (I/O) interface 212, shutter button

30   102, LCD 112, and user input device 114. In one embodiment, memory 208 includes some type of random access memory (RAM) and non-volatile memory,

but can include any known type of memory storage. Control software 210 for controlling processor 206 is stored in memory 208. In operation, when a user presses shutter button 102, processor 206 and shutter controller 204 cause image sensor 200 to capture an image. Image sensor 200 then outputs pixel data

5   representative of the image to processor 206. The pixel data is stored in memory 208, and captured images may be displayed on LCD 112.

      I/O interface 212 is configured to be coupled to a computer or other appropriate electronic device (e.g., a personal digital assistant), for transferring information between the electronic device and camera 100, including

10   downloading captured images from camera 100 to the electronic device.

      Figure 3 is a block diagram illustrating major components of a digital camera 300 configured to perform OCR and text stitching according to one embodiment of the present invention. As shown in Figure 3, digital camera 300 includes the same features as prior art digital camera 200, and also includes OCR

15   and text stitch software 302 stored in memory 208. In one embodiment of the present invention, a plurality of partially overlapping images of a document are captured with camera 300, and OCR is performed on each image by software 302 to convert each image into a text file. Each text file includes a portion of the overall text in the original document. A "text stitch" algorithm is then

20   performed by software 302 to combine the text files. One embodiment of a text stitch algorithm is discussed in further detail below with reference to Figure 5.

      Figure 4 is a diagram illustrating four overlapping image frames 402A-402D of text information from a document 400 captured by camera 300. Three overlap regions 404A-404C are also shown in Figure 4. Overlap region 404A

25   represents the overlap between frames 402A and 402B. Overlap region 404B represents the overlap between frames 402B and 402C. And overlap region 404C represents the overlap between frames 402C and 402D. The processing of images 402A-402D by camera 300 is discussed in further detail below with reference to Figure 5.

30   Figure 5 is a flow diagram illustrating an OCR and text stitching algorithm 500 performed by camera 300 according to one embodiment of the

4

present invention. In step 502, a plurality of partially overlapping images 402A-402D (shown in Figure 4) of a document 400 are captured by camera 300. In one embodiment, camera 300 keeps track of the order in which each image is captured, and stores corresponding order information in memory 208. In step

5   504, camera 300 performs OCR on each of the captured images and generates a text file for each image. Techniques for performing OCR on digital images are known to those of ordinary skill in the art. After camera 300 performs OCR on each one of the individual frames 402A-402D, the text file for frame 402A includes the text "ABCD" and "GH." The text file for frame 402B includes the

10   text "CDEF," "GHIJ," and "LM." The text file for frame 402C includes the text "EF," "IJK," and "LMNO." And the text file for frame 402D includes the text "NOP."

In steps 506 and 508, software 302 compares the text file for the first captured image 402A with the text file for the second captured image 402B, and

15   identifies common characters and/or text strings between the two files. As shown by overlap region 404A in Figure 4, images 402A and 402B (and their corresponding text files) include the characters "CD" and "GH" in common. In step 510, based on the identified common or overlapping text, camera 300 combines the text files for images 402A and 402B into a single text file. The

20   combined text file includes all of the text from the text file for image 402A (i.e., "ABCD" and "GH"), plus all of the non-overlapping text of the text file for image 402B (i.e., "EF," "IJ" and "LM"). The text files for images 402A and 402B are combined by essentially superimposing the text file for image 402B on the text file for image 402A and aligning the common or overlapping text

25   portions. After combining the text files for frames 402A and 402B, camera 300 has recreated the complete first line of text "ABCDEF" from the original document 400.

In step 512, camera 300 determines whether there are any more text files to be processed. If there are no more text files to be processed, the algorithm is

30   done, as indicated at step 514. If there are more text files to be processed, as there are in this example, the algorithm jumps back to step 506. At steps 506

and 508, camera 300 compares the text file for image 402C to the combined text file for images 402A and 402B, and identifies common characters and/or text strings between the two files. As shown by overlap region 404B in Figure 4, images 402B and 402B (and their corresponding text files) include the characters

5 "EF," "IJ" and "LM" in common. In step 510, based on the identified common or overlapping text, camera 300 combines the text file for image 402C and the combined text file for images 402A and 402B into a single text file. The combined text file includes all of the text from the text files for images 402A-402C, with any redundancy from overlapping text portions being eliminated.

10 In step 512, camera 300 once again determines whether there are any more text files to be processed. Since there is one more text file in this example, the algorithm jumps back to step 506. At steps 506 and 508, camera 300 compares the text file for image 402D to the combined text file for images 402A-402C, and identifies common characters and/or text strings between the

15 two files. As shown by overlap region 404C in Figure 4, images 402C and 402D (and their corresponding text files) include the characters "NO" in common. In step 510, based on the identified common or overlapping text, camera 300 combines the text file for image 402D and the combined text file for images 402A-402C into a single text file. The combined text file includes all of the text

20 from the text files for images 402A-402D, with any redundancy from overlapping text portions being eliminated. After combining the text files for all of the image frames 402A-402D, camera 300 has recreated the entire text from the original document 400.

Figure 6A is a diagram illustrating a front side of a combined cellular

25 telephone and digital camera device 600 configured to perform OCR and text stitching according to one embodiment of the present invention. Figure 6B is a diagram illustrating a back side of the combined cellular telephone and digital camera device 600 shown in Figure 6A. Device 600 includes upper portion 600A and lower portion 600B, which may be rotated about hinge 610 to go from

30 an open position (as shown in Figure 6A) to a closed position, as is common with many current cellular telephone models. Device 600 includes antenna 602,

speaker 604, digital camera lens 606, LCD 608, navigation and control buttons 612, and numeric keypad 614. As will be understood by one of ordinary skill in the art, in addition to including lens 606, the digital camera of device 600 also includes conventional internal components, including an image sensor 200,

5   shutter controller 204, processor 206, and memory 208 (shown in Figure 3).

In addition to displaying information regarding cellular telephone operation, LCD 608 is also used as a viewfinder for the digital camera of device 600, and displays captured images. Although no optical viewfinder is shown for device 600, it will be understood by a person of ordinary skill in the art that

10   device 600 could incorporate an optical viewfinder, as well as any other conventional features of currently available digital cameras.

Navigation and control buttons 612 and numeric keypad 614 are used to enter information, navigate through menus displayed on LCD 608 and select menu items, and control operation of device 600. Any one of buttons 612 or 614

15   may be designated as a shutter button 102 for capturing images with the digital camera of device 600, or a dedicated shutter button 102 can be provided.

In one embodiment, device 600 includes a processor 206 (shown in Figure 3) and OCR and text stitch software 302 stored in a memory 208 (also shown in Figure 3) to control operation of the device 600 and to perform OCR

20   and text stitching functions as described above.

In an alternative embodiment of the present invention, rather than performing the OCR and text stitching functions within camera 300 or device 600, images are downloaded from camera 300 or device 600 to a computer or other electronic device, and the computer or other electronic device performs the

25   OCR and text stitching functions described herein.

Although one embodiment of the present invention has been described in the context of a combined cellular telephone/digital camera device, it will be understood by a person of ordinary skill in the art that the techniques disclosed herein are applicable to any device that incorporates a digital camera, including

30   but not limited to, a PDA and a laptop computer.

In addition to allowing a user to manually capture images for OCR and text stitching, in one embodiment, camera 300 and device 600 include a user-selectable automatic capture mode, wherein images are automatically captured at predefined intervals. In the automatic capture mode, the user need only scan the camera over the document to be imaged, and images are automatically captured at equally spaced time intervals.

In one form of the invention, to facilitate the stitching of text files together, a user may input direction information into camera 300 or device 600, which indicates to OCR and text stitch software 302 the direction that images are being captured (e.g., left to right, right to left, top to bottom, bottom to top). In an alternative embodiment, camera 300 and device 600 include a motion sensor for detecting the direction that the device is moving when capturing images, which is used by OCR and text stitching software 302 to facilitate stitching the text files together.

It will be understood by a person of ordinary skill in the art that functions performed by devices 300 and 600, including functions performed by OCR and text stitch software 302, may be implemented in hardware, software, firmware, or any combination thereof. The implementation may be via a microprocessor, programmable logic device, or state machine. Components of the present invention may reside in software on one or more computer-readable mediums. The term computer-readable medium as used herein is defined to include any kind of memory, volatile or non-volatile, such as floppy disks, hard disks, CD-ROMs, flash memory, read-only memory (ROM), and random access memory.

Although specific embodiments have been illustrated and described herein for purposes of description of the preferred embodiment, it will be appreciated by those of ordinary skill in the art that a wide variety of alternate and/or equivalent implementations may be substituted for the specific embodiments shown and described without departing from the scope of the present invention. Those with skill in the chemical, mechanical, electro-mechanical, electrical, and computer arts will readily appreciate that the present invention may be implemented in a very wide variety of embodiments. This

application is intended to cover any adaptations or variations of the preferred embodiments discussed herein. Therefore, it is manifestly intended that this invention be limited only by the claims and the equivalents thereof.